Data Appendix to "From Today's City to Tomorrow's City: An Empirical Investigation of Urban Land Assembly"

NOT FOR PUBLICATION

Leah Brooks
Trachtenberg School of Public Administration and Public Policy
George Washington University

Byron Lutz
Research Division
Federal Reserve Board of Governors

November 2015

This appendix describes how we created the Los Angeles County parcel dataset. Section 1 describes the input datasets. Section 2 describes how we join these datasets, and reports statistics on the uncleaned data. Section 3 describes how we clean the joined dataset, and Section 4 reports statistics on the quality of the final dataset.

1 Data Sources

The basic unit of analysis is the parcel, which is an individual property as legally defined by the Los Angeles County Assessor and Recorder. In any year, there are roughly 2.3 million parcels in the County of Los Angeles. We rely on a number of different sources for information about parcels.

1.1 Parcel-Level Data

For detailed property information on parcels, we rely on data from three separate vendors: DataQuick, Applied Geodetics, and the Los Angeles County Assessor directly.

1.1.1 DataQuick: 1999 - 2002

Dataquick is a property information vendor. It purchases property information from the Los Angeles County Assessor to sell to real estate professionals. We rely on Dataquick data for 1999-2002, and data are reported as of January of each year.

As far as we can ascertain, the Dataquick data are a slightly modified version of the file that the Assessor prepares (specifically, the "Secured Basic File" we discuss below). Dataquick modifies the data from the original Los Angeles County format, and we re-modify it to be consistent with the two following datasets. We discuss modifications at length in Section 3. This data vendor is abbreviated in tables as DQ.

1.1.2 Applied Geodetics: 2004 - 2006

Applied Geodetics is a mapping firm in Los Angeles County. Applied Geodetics sold us data for 2004 (April), 2005 (February) and 2006 (May). To the best of our knowledge, these data are the unmodified Assessor's "Secured Basic File," which is the most complete record of property attributes available to the public from the Assessor. This data vendor is abbreviated in tables as AG.

1.1.3 Los Angeles County Assessor, Local Roll: 2007 - 2009

From 2007 through 2009, we have purchased data directly from the Assessor. Due to financial constraints, we purchased the "Local Roll" database (roughly $400) instead of the "Secured Basic File" (roughly $13,000). The Local Roll has fewer parcel attributes than the Secured Basic File, and comes out annually in July. This data source is abbreviated in tables as LR.

1.1.4 Los Angeles County Assessor, Secured Basic File: 2010 and 2011

In 2010 and 2011, we purchased the Secured Basic File, which is the County's most complete publicly available dataset about properties. These data are from July of each year, and this data source is abbreviated in tables as SB.

1.2 Sales Data

1.2.1 Last Three Sales, 1980 to 2006

In 2006, we purchased a file from the County Assessor that contains information on the last three transactions for each property in the county. For each transaction, we observe transaction type, sale amount (if applicable), and date of transaction.

1.2.2 Sales Within Two Years: 2008, 2009, 2010, and 2011 Files

In 2008, 2009, 2010, and 2011, we purchased additional lists of sales data from the County Assessor. These contain information on all transactions in the prior two years. For each transaction, we observe transaction type, sale amount (if applicable), and date of transaction.

These files leave a small gap from May through December 2006 which we have not been able to fill.

1.3 Parcel Change Database

At our request, the Assessor made a special file that includes all parcel changes from July 1999 to January 2009. Specifically, for each change, this file includes the old parcel number(s), the new parcel number(s), and the effective date of the change. The County has electronic records for parcel changes starting in July 1999 and continuing to the end of our data.

This change database allows us to isolate land assembly and disassembly. The California Assessor's Handbook mentions only one reason for a parcel number to change: if the physical boundaries of a parcel are modified (California State Board of Equalization, 1997, page 26).

We purchased this change database again in July 2009, 2010 and 2011 (each covers all changes in the past two years) to allow us to link all later parcels with previous parcels.

1.4 Digital Parcel Maps

For each year since 2006, we have an electronic map of all parcels that exist in that year. These maps have a boundary (a polygon) for each individual parcel. For each parcel, we use ArcGIS to calculate the x- and y-coordinates (latitude and longitude) of the polygon's geographic center (centroid).

1.5 Census Tract and Block Group Identification

The Census provides census tract and block group boundaries in shapefile format.[1] We use ArcGIS to intersect the 2000 census boundaries with the parcel boundaries from 2006 to 2011 to assign each parcel to a census block group.

The vast majority – 98 percent – of ever-existing parcels have block group identifiers.

1.6 Block Group Data

We use demographic block group level data from the 2000 Decennial Census (ICPSR file 13346, summary level 150), and from the 1990 Decennial Census (ICPSR 9782, summary level 150, but California file is damaged so we used a similar file downloaded from UCLA ATS).

We use ArcGIS and the Census's electronic maps to make a linkage between 1990 and 2000-based block groups, where relationships are based on land area overlap.

1.7 Non-Parcel Digital Maps

- Parks: Information from 2008 ESRI files of local and national parks for California
  - parks displayed on maps are only those more than 1.25 square miles
- Freeways: Data from State of California Cal-Atlas Geospatial Clearinghouse
  - website is http://www.atlas.ca.gov/download.html
  - transportation → Census 2000 → state_highways.* and us_highways.*
- Freeway Entrances and Exits
  - Tele-Atlas US Data, contains federal interstate highway entrances and exits
- Coastline
  - layer of points every 1000 feet along LA County coastline

[1] Files are at http://www.census.gov/geo/www/cob/bdy_files.html.

  - created by taking Census 2000 county map and deleting non-coastline portions
  - used X-tools feature to points to convert coastline line to points
- Metrolink Stations
  - Commuter rail stations
  - File received from Javier Minjares, Southern California Association of Governments, 2010
- Metro Rail Stations
  - Intra-urban rail stations
  - File received from Javier Minjares, Southern California Association of Governments, 2010
- Major Roads
  - Tele-Atlas US Data, version 9.3
  - Major roads only
- Municipal Boundaries
  - Shapefile from 2000 Census

2 Initial Data Linking

Each parcel is identified by a 10-digit number: MMMM-PPP-XXX. The first four digits are the "map book" number – literally the number of the "book" in which the parcel appears. The second three digits are the map book page and identifies the page in the book on which the parcel is found. Each map book page contains a set of geographically contiguous parcels. The last three digits identify individual parcels on the map book page.

We began the data assembly by attempting to link all the annual cross-sections described above (1999-2009, without the missing 2003) using the parcel change database. Panel A of Table 1 presents the results of this original linkage. Slightly fewer than 2 million parcels never change their number throughout the sample (column 2). Column 3 reports the number of ever-changing parcels; this number varies by year as a changer could be 3 parcels in 1999,

but 12 (or 1) parcels in 2009. The number of changes in any given year varies from 35,627 to 157,873. Column 4 reports the number of parcels that exist in this year of the sample, but not in all years of the sample (following a specific definition we describe later). Column 4 shows a striking number of parcels that exist after 2004, but not before. These missing, or "phantom," observations are due to DataQuick editing. We discuss how we deal with these discrepancies in the next section.

We also analyze total land area, measured in square feet, in each category as a check on our data assembly process. We report annual totals in Panel B of Table 1. Columns 2 through 4 report the total amount of parcel square footage in each of the parcel assembly categories by year. Unfortunately, for 2004 and 2005 we do not observe parcel size in the data. It is very clear that across all categories, but particularly in the "phantom" category, that the total land area is drastically smaller in the earlier data source (DataQuick). The California Department of Water Resources measures the total land area of Los Angeles County as 132,487,077,888 square feet, which is clearly much closer to the later data source (143 billion square feet) than the earlier one (22 billion square feet).[2] As we discuss later in great detail, we undertake two major modifications to make total County lot size consistent over time. First, we make sure that lot size for condominiums is coded consistently across all years. Second, we add in parcels that (mostly) do not appear in the DataQuick data that very likely existed at the time of these data (1999 to 2002).

Table 2 shows the total number of parcels (Panel A) and total land area (Panel B) by four categories of use types and by year. The Assessor assigns each property a "use code" that describes how the property is currently used, as distinct from how it is zoned. We use this code to make four major categories of use: single-family, multi-family, vacant, and other. The number of vacant parcels by year (column 4) shows a large break concomitant with the

[2]Parcels cover all land and water area of the county with the exception of public roadways. We discuss later how the latter datasets "overcount" the county's land area.

break in data sources, and similar to the break for the "phantom" parcels of the previous table.[3] The bottom panel of the table shows that the break appears across types.

3 Cleaning and Consistency Changes

This section details the work we did to make the property data consistent across data sources and time.

3.1 Defining the Parcel Group

For ease of analysis, we define a "parcel group" to include all parcels linked by a change. For example, consider a change where parcels A and B combine to C, parcel C splits to D and E, and E splits to F and G. All parcels A through G are in one parcel group, following our definition. We create a unique identifier for each parcel group.

3.2 No Observation of Land Area in 2004 and 2005

The data source we use for 2004 and 2005 does not contain information on parcel land area. We assume that all 2004 parcels existing in 2002 have the same land area that they did in 2002, and all 2005 parcels existing in 2006 have the same land area they will have in 2006. These are very reasonable assumptions, as lot size changes only when parcel number does. As we describe in greater detail below, we use a variety of methods to evaluate and clean lot size across time.

If a 2004 or 2005 parcel is also missing lot size information in 2002 or 2006, we take lot size from the closest year for which it is available.

[3] Vacant parcels may be coded vacant residential, vacant commercial, or simply vacant. "Vacant" appears to refer to habitation or use, and not to the presence of structures.

3.3 Vertically Stacked Parcels

The assessor draws the boundary for a single piece of land with multiple owners, such as condominiums, so that each owner's parcel has the land area of the entire lot. This type of separate ownership of the same piece of land is distinct from joint ownership of a single property. Separate ownership, or condominium-type ownership, parcels are "stacked" vertically on the Assessor's map, in the same way you would stack checkers. In this case, think of each checker as the condominium of an individual owner. A 100-unit condominium will consist of 100 parcels, one for each legally distinct unit of ownership. Stacked parcels are usually, but not always, condominiums.

In the DataQuick data, the land area of each condominium unit is the land area of the lot divided by the number of parcels covering that lot. In the remaining datasets, when units and parcels are co-terminus, each unit is assigned the land area of the entire lot. This causes problems with multifamily parcels in particular, and is one of the two causes of the discrepancy in land area between the 1999 to 2002 and 2003 to 2010 periods in Tables 1 and 2. We correct for this problem by modifying all parcels to have a land area consistent with the method in the DataQuick data (land area of a parcel = total land area of parcel / number of parcels occupying that lot). We prefer this method of measuring land area because the net land area of the county remains unchanged even if one parcel becomes three condos. The method used in the later datasets suggests that the total land area of the county increases when condominiums are created.

We identify vertically stacked parcels by calculating the centroid of every parcel on each map we have (annual maps, 2006-2010). In any given year, all parcels that share the same x- and y-coordinates are "stacked," and their land area requires correction. On average, a map-year has 251,518 unique parcels that are stacked, 15,299 stacks, and a mean of 16 parcels per stack. Unfortunately, we do not have maps for years before 2006, so this correction is incomplete for 2004 and 2005. Specifically, we can correct all stacked parcels that exist in

2004 and 2005 and which continue to exist in 2006. If parcels exist in 2004 and 2005 only – with no 2006 data – we are not able to re-calculate a lot size for those parcels. (Section 3.6 addresses how we handle observations which we are missing lot size information.)

3.4 "Phantom" Parcels

The third data challenge we faced was the presence of parcels that did not appear in all years, and the disappearance of which could not be explained by any observation in the parcel change database. We call these appearing and disappearing parcels "phantoms." Specifically, to be a "phantom," in our terms, a parcel must appear in at least five years of data.

Table 1 suggests that the DataQuick data do not include all parcels that exist in the county. We strongly suspect that, in an effort to satisfy customers primarily interested in residential (or at least tradeable) property, DataQuick deletes parcels that will not be transacted.

Table 3 describes the appearance of phantoms across the years of the final dataset. The horizontal axis has one column for each year of the dataset. The vertical axis has one row for each year of phantom non-appearance. The (1999,2000) cell says that 649 parcels that exist in the 1999 dataset do not exist in the 2000 dataset and are not accounted for by any parcel changes. The diagonal axis is by definition zero, since a parcel cannot be missing in the year in which it exists. The clear and striking pattern in this table is that the vast majority of phantom parcels are missing in the years 1999 to 2002, and come from datasets in years 2004 to 2010.

To resolve this problem, we say that any parcel that exists for at least five years in the dataset is a "true" parcel. Because these parcels are not in the change database, we assume that their size is constant over the period of interest and that they exist in all years. Roughly 200,000 of these parcels truly exist from 2004 to 2010 and do not appear from 1999 to 2002.

Only about 20,000 parcels appear from 1999 to 2002 and not from 2004 to 2010. Were we not to make this adjustment, we would have an severely unbalanced panel, and the early years of the panel would not account for a substantial amount of the land area of the County.

3.5 Relating Parcel Change to Parcel Cross-sectional Databases

We use the assessor's parcel change database to link parcels across years. The assessor reports on changes of four types: one to many, one to one, many to one, zero to one and one to zero. These last two changes (zero to one and one to zero) are the least frequent of all change types. They are used to move land in and out of the county. Land in use by roads is not in parcels, so the creation of a parcel from a road could cause a zero to one change. Similarly, a one to zero change could occur when a parcel becomes part of a road.

Table 4 reports on the quality of the match between the datasets. The table reports statistics for all changing parcels, and by change type. Columns 2 and 3 show that more parcels enter the county than leave over the period. Columns 4 and 5 report the number of parcels from the change database that find a match in the cross-sectional database, and shows whether these parcels are entering or leaving the county.

Column 8 reports the share of parcels from the change database that we consider "successfully matched." Across all years, we are able to match 97 percent of parcels in the change database with parcels in the cross-sectional databases. Matches can be one of two types. The first type of match is when a parcel in the change database directly matches a parcel in the cross-section. These are a majority of the matches, as shown in Column 6. However, many-to-many changes (and some others) require the use of placeholder parcels; this is the second type of match. The Assessor's file structure does not allow for a direct many-to-many match. Instead, many-to-many matches are coded as many-to-one and then one-to-many. In this case, the "one" is the placeholder parcel. We report the share of matches to placeholder parcels in Column 7. Across all types of changes, roughly 17 percent of changing parcels

match to a temporary placeholder parcel. As shown in Column 9, only three percent of parcels have no match at all. The final column in the table reports the time from appearance in the parcel change database to appearance in the cross-sectional database. Parcels usually take about one year to appear in the cross-section database after appearing in the change database.

To date to a parcel change, then, we must make a choice between two possible dates – one from the cross-sectional data, and one from the parcel change database. In cases where the parcel change database and the cross-sectional databases do not agree, we defer to the change database. The change database records an administrative, or policy change, while the cross-sectional data more likely reflects what is actually physically changing. Since we are most interested in policy choices regarding land use, the date of the decision to change a parcel seemed more relevant than the physical change itself.

Finally, we drop a very small number of parcels we cannot link to other parcels in the cross section to form a change group. Specifically, we drop 4,812 observations (470 unique parcel numbers) that do not fit into a change group.

3.6 Reported Changes in Land Area for Non-Changing Parcels

We check the internal consistency of the data by examining whether there are changes in land area for parcels that do not change parcel number across the years of the sample. By construction, such parcels should have the same land area for all years of the sample. For each parcel, we calculate the percentage difference between the largest land area and the smallest land area reported (relative to the mean lot size of the entire period). Though the median change is a very small 0.7 percent, changes at the ninetieth percentile and above can be quite large.

Given this, we make two adjustments to clean the land area variable. First, if a parcel's land area changes by less than 20 percent from 1999 to 2011, we give the parcel the 2011

land area value for each year. We rely on data from the later years of the sample because by all measures we report here it is more reliable than the DataQuick data. This change modifies the values for virtually all of the parcels in the dataset. Of the modified parcels, about 19 percent of them have very minor differences of less than 0.1 percent, which seem to be due to rounding differences between the different datasets. Another 60 percent have differences between 0.1 and 5 percent of land area over time, and 6 percent have differences of between 5 and 20 percent.

The second adjustment to clean the land area variable modifies parcels that are vertically stacked. After correcting for the different lot size definitions, if the percentage change in lot size for the same parcel is less than 20 percent, we replace the area for all years with the average across all years in the sample.

Finally, we delete 74,787 unique parcels, between 2 and 4 percent of the observations in each cross-section, that have land area changes of more than 20 percent, or which have always-missing values for land area.

3.7 Land Area Adjustments for Parcel Changes

When parcels change, the total initial lot size of changing parcels should equal the final total lot size of the changing parcels. Unfortunately, this is not always the case.

For example, if a single parcel changes number and becomes only one new parcel, the parcel's geographic area does not change by construction. However, due to dataset breaks and other inconsistencies, differences in lot size sometimes appear. These create inconsistencies in geographical areas over time that do not reflect actual changes to parcels.

We resolve this problem by inflating or deflating each parcel's lot size so that the total lot size for parcels that change within a parcel group is the same for each year of data. Let g be the parcel group, which is all parcels ever associated with a given change. The adjustment is

$$\text{lot size}_{i,g,t} = \text{reported lot size}_{i,t} * \frac{\text{average total lot size}_g}{\text{total lot size}_{g,t}}.$$

The final value for parcel i in year t ($\text{lot size}_{i,t}$) is equal to parcel i's reported square footage in year t ($\text{reported lot size}_{i,t}$) multiplied by the ratio of the average of the parcel group's total square feet ($\text{average total lot size}_g$) divided by the current year total square feet ($\text{total lot size}_{g,t}$). By adjusting at the parcel group level, this method allows us to account for both simple and complex changes.

The result of this adjustment is that each parcel group has a constant lot square footage across time. This better reflects the reality of a parcel change. Square footage at the beginning should equal square footage at the end, regardless of how the land is divided or combined.

3.8 Missing Values for Parcel Attributes

In addition to the work with lot square feet that we do above, we fill in missing values for parcel attributes to a limited extent. For non-geographic information, we edit parcels that do not change number, and for which a later observation of the parcel exists. In these cases, we attribute the later parcel characteristics to the earlier parcel if the structure on the parcel was built before the year of the data. For example, imagine parcel A exists from 1999 to 2009, but we observe attributes starting only in 2005. If the structure on parcel A was built in 1995, we fill in the missing values for parcel A from 1999 through 2004 with the 2005 values. If the structure on parcel A was built in 2005, we do not fill in missing values (save for lot square footage as described above).

Geographic identifying information – tract and block group information, and latitude and longitude information – come from merging the panel data with information generated from electronic maps. We have these electronic maps starting only in 2006. For parcels that

do not change, we know that the physical border has not changed and therefore use the 2006 geographic information.

For parcels that existed before 2006 and not afterward we impute geographic information. Specifically, we assign old parcels to the same block group or tract as the later-existing parcels (no parcel change groups cross block groups). For the latitude and longitude, we give the old parcels the average of the latitude and longitude of the later-existing parcels. This should be a very close approximation to the actual location.

For parcels that are not assigned a block group through these methods (changing or not), we interpolate to find the block group. Specifically, if a parcel is missing a block group, we assign the parcel to have the block group of all other parcels on that map book page. A map book page is identifiable by the first seven digits of the parcel number. Map book pages have very few parcels and very rarely cross block group boundaries.

3.9 Determining Owner Occupancy Status and Housing Tenure

In order to determine whether a parcel is owner-occupied, we compare the parcel's physical address to the parcel's mailing address used for tax purposes (i.e., where the property tax bill is sent). If the two addresses match, we infer that the property is owner-occupied. In practice this procedure is complicated by the fact that the address data is often missing or incomplete. We therefore use an algorithm to classify each parcel into one of three categories: match, non-match, and insufficient data.

The address fields in our data are state, zip code, city name, street name and street number. We categorize a parcel as a match, and hence as owner-occupied, if the street number is not missing *and* either of the following two conditions pertaining to the mailing and physical addressed are met:

1. The street number in the first N places is the same in the physical and mailing address, where N is the length of the shorter of the two street numbers (if $N < 3$ then the length

of the two street numbers must also be equal), *and* the first three letters of the street name are the same in both the physical and mailing address.

2. The same street number in the first N places is the same in the physical and mailing address (if $N < 3$ then the length of the two street numbers must also be equal) *and* the physical and mailing addresses are equal for any two of the following :

    - First letter of the street name
    - First letter of city name
    - First letter of state name
    - First digit of the zip code

We categorize a parcel as a non-match, and hence as not owner-occupied, if it is not categorized as a match *and* any one of the following conditions for the mailing and physical addresses are met:

- Both state names are valid and differ *and* one of the following:
    - First three letters of city name are not identical (and neither is missing)
    - First three letters of street name are not identical (and neither is missing)
    - First two digits of zip code are not identical (and neither is missing)
    - First two digits of street number are not identical (and neither is missing)
- Mailing address is a PO Box
- Street numbers differ in the first N places and are not missing
- Both zip codes are valid, differ *and* one of the following:
    - First three letters of city name are not identical (and neither is missing)
    - First three letters of street name are not identical (and neither is missing)
    - First two digits of street number are not identical (and neither is missing)
- The first four characters of the street name do not match *and* the first four characters of the city name do not match
- Street name matches exactly and the street number does not

We categorize a parcel as having insufficient data if it is neither a match nor a non-match.

Among the subset of parcels identified as being owner-occupied, we infer housing tenure from the number of years since the house last sold. For instance, an owner-occupied single family house which last sold in 2000 is categorized as having housing tenure equal to 5 in 2005.

3.10 Determining Which Assemblies are Teardowns

For a given assembly, pre-assembly, we observe attributes for one structure per parcel. However, a parcel may, in fact, contain more than one structure. For each parcel change group before assembly, we find the minimum and maximum of the age of the structure on each parcel: $\max(age_{g,before})$ and $\min(age_{g,before})$ (g indicates a parcel change group). We observe similar ages post-assembly: $\max(age_{g,after})$ and $\min(age_{g,after})$. We define parcels as being assembly teardowns only if $\max(age_{g,before}) < \min(age_{g,after})$. This definition therefore excludes change groups where some, but not all, structures were demolished.

3.11 Preparing Sales Data

3.11.1 Basic Cleaning

To prepare the sales data, we restrict our analysis to arm's length transactions. We drop sales with dates not on any calendar (e.g., day 45; this is a problem only for the 2006 dataset), sales with values of 10 or less (which the assessor uses as a special code for adjoining pools or structures), and sale amounts equal to 999,999,999. We drop sales before 1985. We keep only one observation if the data report two sales on the same day of the same amount. A small share of observations have two sales on the same day of differing amounts. In this case, we drop both observations. Finally, we drop all sales with a value of less than \$0.15 per square foot, as we believe that this value is too small to be credible. This restriction drops less than 1 percent of the remaining sales.

3.11.2 Multi-sales

In some cases, parcels do not sell individually. For example, suppose that three parcels sell as a group at one time. In such a case, the assessor observes only the total purchase price. In these cases, the Assessor codes the sale as a "multi-sale." Multi-sales are 6 percent of all sales.

The Assessor codes multi-sales as 2 through 9 or M (more than 9) to designate the number of parcels in the multi-sale. In the raw data, the assessor puts the total price as the price for each parcel, therefore overstating total sale amounts.

We correct for this overstatement by re-scaling prices. To begin, we double-check the assessor's coding of multi-sales. We count the number of sales that occur on the same day, on the same map book page, for the same sale amount. Any sales that are duplicates by these terms we designate as multi-sales. If we count the same number of parcels in the multi-sale as the Assessor reports, we call this a verified multi-sale. Our analysis uses only verified multi-sales and drops non-verified multi-sales.

We correct sale values for multi-sale properties by giving each property a share of the total sale price equal to its share of the lot square footage of all the properties in the group.

4 Cleaning and Consistency Results

Tables 5 and 6 present the results of the data cleaning. In the first of these tables, Panel A shows that the overall pattern of the number of parcels over time no longer exhibits any breaks when we change data sources. The number of phantom parcels and never-changing parcels is now consistent across all years. The number of changing parcels increases over time: there are roughly 60,000 of these parcels in 1999 and 160,000 of them in 2011. This means that, on net, more parcels have split than merged. Panel B shows that this increase is not driven by substantial changes in land area. This table no longer shows discontinuous

patterns across dataset breaks.

Panel A of Table 6 shows that, from 1999 to 2011, the number of single- and multi-family and vacant parcels increased, while the number of "other" parcels decreased. Land area trends, shown in Panel B, also no longer show sharp discontinuities related to dataset breaks.

References

California State Board of Equalization, 1997. "Assessor's Handbook 215: Assessment Map Standards." Tech. rep., California State Board of Equalization.

Table 1: Original Dataset: By Type of Appearance in Dataset

		(1)	(2)	(3)	(4)
		Panel A: Parcels			
Year	Source	Total	Never-Changers	Changers	Phantoms
1999	DQ	2,054,630	1,948,502	35,627	6,187
2000	DQ	2,062,666	1,948,502	38,239	11,294
2001	DQ	2,087,855	1,948,502	46,323	26,114
2002	DQ	2,087,417	1,948,502	54,704	17,449
2004	AG	2,311,387	1,948,502	86,692	206,481
2005	AG	2,319,526	1,948,502	94,412	207,219
2006	AG	2,337,233	1,948,502	111,539	207,947
2007	LR	2,354,414	1,948,502	128,751	208,400
2008	LR	2,366,832	1,948,502	140,705	208,935
2009	LR	2,376,360	1,948,502	150,182	209,330
2010	SB	2,380,386	1,948,502	154,261	209,647
2011	SB	2,383,988	1,948,502	157,873	209,677
		Panel B: Land Area, Square Feet			
Year	Source	Total	Never-Changers	Changers	Phantoms
1999	DQ	27,764,230,819	19,504,676,079	1,124,038,343	418,477,358
2000	DQ	22,403,685,543	19,505,037,945	1,011,368,413	546,960,947
2001	DQ	23,929,788,784	19,505,038,780	1,224,775,567	1,668,968,971
2002	DQ	22,411,729,309	19,505,073,265	1,111,671,740	403,388,448
2004	AG				
2005	AG				
2006	AG	138,025,141,417	45,151,780,276	9,510,647,933	72,011,080,923
2007	LR	141,249,361,785	45,154,323,925	10,296,129,745	72,162,588,150
2008	LR	142,126,764,819	45,153,989,843	11,197,350,796	72,244,152,453
2009	LR	142,791,412,096	45,153,547,001	11,765,966,573	72,389,504,481
2010	SB	143,083,855,131	45,153,039,051	12,139,840,358	72,466,658,923
2011	SB	143,215,366,799	45,131,956,782	12,500,698,840	72,345,117,341

Note: Columns 2, 3 and 4 do not add up to the total in Column 1. Column 1 reports the total number of parcels in the original data. As described in this appendix, we drop some parcels. These dropped parcels are not classified as being one of the three types in Columns 2, 3 and 4.

Table 2: Original Dataset: Use Type By Year

		(1)	(2)	(3)	(4)	(5)
			Panel A: Parcels			
Year	Source	Total	Single-Family	Multi-Family	Vacant	Other
1999	DQ	2,054,630	1,399,650	436,693	16,332	127,788
2000	DQ	2,062,666	1,400,390	436,686	16,448	127,915
2001	DQ	2,087,855	1,400,729	436,658	28,288	127,898
2002	DQ	2,087,417	1,401,260	436,982	15,780	127,509
2004	AG	2,311,387	1,435,886	455,012	184,643	163,087
2005	AG	2,319,526	1,438,901	459,453	184,088	163,251
2006	AG	2,337,233	1,448,345	469,750	182,511	163,790
2007	LR	2,354,414	1,455,306	480,489	181,534	163,177
2008	LR	2,366,832	1,458,101	492,126	179,772	164,284
2009	LR	2,376,360	1,459,853	501,540	178,657	164,334
2010	SB	2,380,386	1,460,179	507,560	177,099	164,267
2011	SB	2,383,988	1,461,338	511,307	176,493	163,971
			Panel B: Land Area, Square Feet			
Year	Source	Total	Single-Family	Multi-Family	Vacant	Other
1999	DQ	27,764,230,819	13,720,572,239	2,692,490,776	663,269,634	3,970,844,134
2000	DQ	22,403,685,543	13,730,266,220	2,690,080,642	662,540,286	3,940,631,065
2001	DQ	23,929,788,784	13,759,487,498	2,688,552,227	1,849,059,181	3,949,719,842
2002	DQ	22,411,729,309	13,775,928,500	2,689,535,220	398,909,792	3,902,599,372
2004	AG					
2005	AG					
2006	AG	138,025,141,417	18,963,069,965	27,976,927,390	40,656,243,630	38,946,710,876
2007	LR	141,249,361,785	19,175,924,780	28,711,136,882	40,648,767,481	39,015,143,144
2008	LR	142,126,764,819	19,290,606,188	29,473,535,434	40,622,373,977	39,161,389,659
2009	LR	142,791,412,096	19,375,483,660	30,192,800,865	40,494,279,725	39,239,717,086
2010	SB	143,083,855,131	19,397,286,430	30,464,889,679	40,624,729,739	39,265,223,833
2011	SB	143,215,366,799	19,447,253,562	30,642,750,267	40,691,087,395	39,196,681,739

Note: Columns 2 through 5 do not add up to the total in column 1, as a very few parcels have no intelligible use codes.

Table 3: Phantom Parcels Across Years

	Total # of parcels that do not appear in vertical axis year, but do appear in this year & for at least five years											
	1999	2000	2001	2002	2004	2005	2006	2007	2008	2009	2010	2011
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1999	0	7,370	25,193	20,302	223,624	222,504	219,819	218,016	215,840	214,836	213,393	212,424
2000	649	0	20,766	17,617	222,718	221,518	218,834	217,033	214,830	213,830	212,391	211,405
2001	1,034	425	0	11,623	209,179	208,135	205,805	204,202	202,147	201,239	199,859	198,945
2002	5,406	4,962	16,085	0	220,580	219,347	216,788	215,007	212,863	211,873	210,423	209,403
2004	2,048	1,465	1,200	1,070	0	6,924	8,326	8,934	9,601	10,219	10,561	10,793
2005	2,054	1,473	1,208	1,078	37	0	12,621	13,231	13,872	14,510	14,822	15,022
2006	2,058	1,477	1,213	1,083	119	94	0	4,475	5,269	5,946	6,329	6,580
2007	2,063	1,482	1,218	1,089	159	136	58	0	2,225	2,967	3,400	3,703
2008	2,063	1,482	1,218	1,089	188	168	96	51	0	3,337	3,825	4,129
2009	2,064	1,483	1,219	1,090	210	190	119	76	36	0	1,455	1,916
2010	2,065	1,484	1,220	1,091	225	207	140	103	69	44	0	1,372
2011	4,171	3,600	3,422	3,253	3,551	3,537	3,475	3,458	3,431	3,410	3,400	0

Notes: This table reports a count of only phantom parcels. Each cell reports the number of parcels in the year of the horizontal axis that are "phantoms" that we have added in the year of the vertical axis. For example, cell (1999,2004) has a value of 2,048 which means that 2,048 parcels exist in 1999 and at least four other years and do not appear in the 2004 cross-sectional data. This table accounts for changes from the change database. By construction, the diagonal is zero – a parcel that we add to the database because it does not appear (a phantom) cannot exist in the original data in that year.

Table 4: Parcels in Change Database by Type of Appearance

	Parcels in Change Database		Parcels in the Cross-Section with match in parcel change database		Share of all parcels in Change Database				
Change type	Entering County	Leaving County	Entering County	Leaving County	Match in Cross-sectional database	Temporary parcel is not matched, but change parcel is	Successfully Matched Parcels	No match in Cross-sectional database	Matches: Average years for change to appear in cross-section *
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
All types	204,832	99,748	178,963	64,137	0.80	0.17	0.97	0.03	1.11
Zero to One	2,400	1	661		0.28	0.62	0.90	0.10	1.49
One to Zero	1	918		458	0.50	0.00	0.50	0.50	
One to One	32,295	32,338	26,535	23,770	0.78	0.20	0.97	0.03	1.18
One to Many	152,914	18,025	139,377	9,456	0.87	0.11	0.98	0.02	1.08
Many to One	17,222	48,466	12,390	30,453	0.65	0.30	0.95	0.05	1.15

* This figure is calculated using only cases where the time for the parcel to appear in the cross-section is positive. In a very small set of cases, a parcel may appear in the cross-sectional data before it appears in the change database, in which case the time for the parcel to appear is negative. Approximately 6 percent of matched parcels have a negative time to appear; 83 percent of these negative values appear in the first year of our sample, 1999.

Table 5: Cleaned Dataset: By Type of Appearance in Dataset

		(1)	(2)	(3)	(4)
		Panel A: Parcels			
Year	Source	Total	Never-Changers	Changers	Phantoms
1999	DQ	2,220,095	1,948,502	59,214	212,379
2000	DQ	2,226,745	1,948,502	65,864	212,379
2001	DQ	2,232,094	1,948,502	71,213	212,379
2002	DQ	2,239,496	1,948,502	78,615	212,379
2004	AG	2,252,467	1,948,502	91,586	212,379
2005	AG	2,263,834	1,948,502	102,953	212,379
2006	AG	2,275,592	1,948,502	114,711	212,379
2007	LR	2,291,112	1,948,502	130,231	212,379
2008	LR	2,304,070	1,948,502	143,189	212,379
2009	LR	2,312,035	1,948,502	151,154	212,379
2010	SB	2,316,179	1,948,502	155,298	212,379
2011	SB	2,321,904	1,948,502	161,023	212,379
		Panel B: Land Area, Square Feet			
Year	Source	Total	Never-Changers	Changers	Phantoms
1999	DQ	101,045,575,947	19,953,151,787	8,676,099,862	72,416,324,298
2000	DQ	101,043,896,594	19,953,151,787	8,674,420,509	72,416,324,298
2001	DQ	101,027,910,070	19,953,151,787	8,658,433,985	72,416,324,298
2002	DQ	101,035,449,637	19,953,151,787	8,665,973,552	72,416,324,298
2004	AG	101,055,095,260	19,953,151,787	8,685,619,175	72,416,324,298
2005	AG	101,042,116,217	19,953,151,787	8,672,640,132	72,416,324,298
2006	AG	101,039,254,204	19,953,151,787	8,669,778,119	72,416,324,298
2007	LR	101,041,168,637	19,953,151,787	8,671,692,552	72,416,324,298
2008	LR	101,051,694,982	19,953,151,787	8,682,218,897	72,416,324,298
2009	LR	101,055,022,345	19,953,151,787	8,685,546,260	72,416,324,298
2010	SB	101,080,139,060	19,953,151,787	8,710,662,975	72,416,324,298
2011	SB	101,319,598,704	19,953,151,787	8,950,122,619	72,416,324,298

Table 6: Cleaned Dataset: Use Type By Year

		(1)	(2)	(3)	(4)	(5)
			Panel A: Parcels			
Year	Source	Total	Single-Family	Multi-Family	Vacant	Other
1999	DQ	2,220,095	1,418,218	445,884	175,673	179,318
2000	DQ	2,226,745	1,422,309	448,362	176,446	178,424
2001	DQ	2,232,094	1,426,138	449,673	178,323	176,744
2002	DQ	2,239,496	1,431,767	451,753	177,975	176,840
2004	AG	2,252,467	1,438,890	458,633	188,018	165,919
2005	AG	2,263,834	1,443,309	465,230	188,135	166,199
2006	AG	2,275,592	1,450,579	473,405	184,990	165,757
2007	LR	2,291,112	1,457,071	483,336	184,411	165,456
2008	LR	2,304,070	1,459,917	495,133	182,062	166,137
2009	LR	2,312,035	1,461,376	503,333	180,483	166,048
2010	SB	2,316,179	1,461,672	509,142	178,578	165,992
2011	SB	2,321,904	1,463,549	512,388	178,865	166,307
			Panel B: Land Area, Square Feet			
Year	Source	Total	Single-Family	Multi-Family	Vacant	Other
1999	DQ	101,045,575,947	15,181,544,607	3,262,746,685	42,464,887,994	39,940,194,500
2000	DQ	101,043,896,594	15,604,921,648	3,290,852,310	42,161,273,648	39,627,551,166
2001	DQ	101,027,910,070	16,672,977,896	3,367,821,918	41,039,886,406	39,609,467,767
2002	DQ	101,035,449,637	16,521,335,291	3,673,878,015	40,945,640,746	39,698,352,668
2004	AG	101,055,095,260	16,776,334,754	3,702,529,345	41,432,839,634	39,101,202,122
2005	AG	101,042,116,217	16,852,988,610	3,699,048,520	41,330,749,509	39,128,076,071
2006	AG	101,039,254,204	16,855,731,407	3,768,149,464	41,267,359,016	39,129,931,492
2007	LR	101,041,168,637	16,975,273,672	3,892,363,301	41,006,052,815	39,163,785,952
2008	LR	101,051,694,982	17,045,615,339	3,919,442,451	40,889,968,790	39,193,243,887
2009	LR	101,055,022,345	17,067,792,621	3,953,296,032	40,775,960,747	39,257,972,945
2010	SB	101,080,139,060	17,079,013,486	3,972,852,109	40,772,913,715	39,255,359,749
2011	SB	101,319,598,704	17,112,122,808	3,985,296,710	40,903,686,375	39,318,492,811

Note: Columns 2 through 5 do not add up to the total in column 1, as a very few parcels do not have an intelligible use code.